DOCUMENT RESUME

ED 463 527 CS 510 852

AUTHOR Rhodes-Kline, Anne K.
TITLE Sustained Effects of Reading Recovery: A Review of the Literature.
INSTITUTION Maine Univ., Orono. Center for Early Literacy.
PUB DATE 1997-09-00
NOTE 17p.; A Center for Early Literacy publication.
PUB TYPE Information Analyses (070)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Early Intervention; Grade 1; *High Risk Students; *Literacy; Literature Reviews; Primary Education; *Program Effectiveness; *Reading Programs; Research Problems
IDENTIFIERS *Maine; Reading Recovery Projects; Research Priorities; Sustaining Effects Study

ABSTRACT

Reading Recovery (RR), an early intervention for first graders at risk of literacy failure, was introduced in Maine in school year 1990-91 and has been growing ever since. The Maine Department of Education has been funding the training of RR teachers since the 1994-95 school year. Although there is general agreement that the RR program is successful at bringing at-risk first graders up to the level of their classmates, a larger question is: Do RR children sustain these gains beyond the exit of first grade? A literature review covered the research as completely as possible, examining some of the studies through 2 years of program evaluation research at the University of Maine and searching the ERIC database. Some of the studies reviewed were conducted at Chicago (Illinois), Newark (New Jersey), New York University, Wake County (North Carolina), Ohio State University, Westbrook (Maine), Texas Women's University, and the University of Melbourne (Australia). Future research into the long-term impact of the RR program is needed. Many of the studies reviewed suffered from serious methodological flaws. For example, ceiling effects were a problem for studies that relied on the text reading-level measure beyond the first grade, while other studies included only discontinued RR children. Also, implementation levels differ so much from school to school, but this information is rarely reported. A longitudinal study is planned for Maine. (Contains 13 references.) (NKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 527

Sustained Effects of Reading Recovery:

A Review of the Literature

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P F Meare

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

September, 1997

Anne K. Rhodes-Kline
Center for Early Literacy
College of Education and Human Development
University of Maine

CS 510 852

BEST COPY AVAILABLE

Sustained Effects of Reading Recovery:
A Review of the Literature

Reading Recovery (RR; Clay, 1982; 1985; 1991) is an early intervention for first graders at risk of literacy failure. It selects the entering first grade children whose literacy skills are the lowest in the class. A specially trained RR teacher works one-on-one with each child for 30 minutes a day, every day that school is in session, to accelerate the children to the average level of the first grade class. Children whose skills levels are successfully raised to the level at which they can continue to progress with only regular classroom instruction are *discontinued* from the program. As soon as one child discontinues, another child who is the next neediest is started in his or her place.

In 1984, the success of a pilot project in New Zealand led researchers at the Ohio State University to introduce RR to the United States. RR was introduced in Maine in school year 1990-91, and it has been growing every year since then. The Maine Department of Education has been funding the training of RR teachers since the 1994-95 school year.

Although there is general agreement from many researchers that the RR program is successful at bringing at-risk first grade children up to the average level of their classmates, a larger question is, do RR children sustain these gains (relative to their classmates) beyond the end of first grade? A number of studies have attempted to answer this question. This paper summarizes these efforts.

Studies were chosen for this report in an effort to cover the research as completely as possible. The author came across some of the studies through the course of two years of RR program evaluation research at the University of Maine. To find others, a search was conducted

of the ERIC database using the search term "Reading Recovery" and any of the following additional terms: "longitudinal," "progress," "follow-up," or "follow up." Studies were included if they examined the progress beyond grade one of students who had received RR.

Ohio State University

In two reports (Pinnell, DeFord, & Lyons, 1988; Ohio State University, 1989), the findings from the first published follow-up study of RR in The United States were described. This study was conducted in Columbus, Ohio (the pilot site for RR in this country) by the educators who brought the program to the U.S. and were responsible for its administration. It was a small to medium sized study, involving over 100 RR children , and 100-150 comparison children in each of its four years.

In this study, RR was compared to other (existing) services for first graders at risk of failing to learn to read. Children who participated in the RR program were compared to children who had received another intervention (i.e., whatever program the school had in place). In schools that had already agreed to implement RR, comparison children were taken from a waiting list of children who could not be served by RR because of limited resources. Because RR selects the lowest skilled children first, these waiting list (WL) children had skill levels somewhat higher than those of RR children, although both groups of children were judged to be at risk. In schools that agreed to participate in the study but that had not officially implemented the RR program, children were alternately assigned to RR or another intervention.

The literacy progress of RR children was compared to a random sample (RS) drawn from the RR children's grade mates in grades 1, 2, 3, and 4. Literacy progress was measured using levels 1 through 30 of the text reading component of the Observation Survey (Clay, 1993). This

assessment involves having a child read an unfamiliar text out loud to an adult listener, while the adult records detailed information about any errors the child makes. The highest level text the child reads with 90% accuracy is the score the child receives. For the final year of the study (4th grade year), the Woodcock Reading Mastery test was also used.

The study was not technically a longitudinal study, in that only RR and WL children were the same for all four years. A new RS was drawn in the spring of each year. This seriously limited the analyses that could be conducted, and it changed the nature of the comparison group, since some groups of children were systematically excluded. For example, children who were retained would not have an opportunity to be chosen for the RS because they would not have been in the grade level from which the RS was drawn. Also, children moving into the district might have been selected for the RS even though they may have been identified as at-risk when they were in first grade. There may have even been some children in the RS who had received RR at a different school.

Another major limitation to the study is that only children who had 60 lessons or more, or who successfully discontinued from the program were included as RR children. No information was given about the other children who received RR. In addition, RR and WL children who had been retained were dropped from some of the analyses. This procedure eliminated 48 out of 112 RR children and 19 out of 39 WL children by the end of 4th grade.

Another limitation to this study was ceiling effects on the text reading level assessment. Ceiling effects are a problem when researchers do not choose measures on which children can display their full abilities. High-skill children "hit the ceiling" when they achieve the maximum possible score. The problem with this in a study which aims to show that two groups are not

different is that the two groups may indeed be different (i.e. RS children may be significantly farther ahead in literacy than RR children) but the higher group could not show this because of limitations with the measures.

Results indicated that, on average, discontinued RR children maintained their gains relative to comparison children. Average scores for discontinued RR children were consistently ahead of at-risk comparison children. Discontinued RR children were, on average, consistently behind RS children, on average, at the end of second, third, and fourth grades. However, the average score for RR children was within the "average band" (the mean of RS scores plus and minus ½ standard deviation) at the end of second and third grades, but just below the average band at the end of fourth grade.

Despite its limitations, this study was an important early attempt to validate the claim that RR's effects on the children who participate are lasting. The study found that, at least among children who are successful in the RR program, gains are sustained relative to those of comparison children, and gains may be sustained relative to RS children.

Chicago, Illinois

A small study (60 third graders) was conducted in a Chicago public school (Curtin, 1993). The researcher administered the Iowa Test of Basic Skills in the spring of both first and third grades. She subtracted the latter from the former to arrive at a gain score for each child. She found that students tutored in RR did not differ from their peers who did not receive RR, and she therefore concluded that RR children sustain their gains into the third grade.

Although a laudable effort, the report describing this study failed to offer some very important details. No information was provided about the comparison children. Were they RS?

Or were they judged to be at risk? The article seems to imply that the comparison children were the not-at-risk peers of RR children. However, this is not specified.

Also, the use of gain scores is not the best way to answer the question of whether RR children sustain their gains. It is not clear from this study whether one group's scores are actually higher or lower than the other group's scores; all that is indicated is that the margin has not changed.

Newark, New Jersey

Ramaswami (1994) conducted another small study to validate the sustained-effects claim of the RR program at the local level. All children who had been served by RR (both discontinued and not discontinued) from 4 schools in the Newark area were included. Seventy-two (72) children were included in all, 24 RR, 30 RS, and 18 WL.

Two notable strengths of this study differentiate it from its predecessors. First, all children who participated in the RR program, regardless of how long or how successfully, were included in the data. Separate analyses could ask questions about discontinued versus not discontinued children. Second, the pool of comparison children was not re-drawn for the study, but was maintained from year to year, making this a true longitudinal study.

The Stanford Achievement Test at the end of second grade was the assessment used. Mean ranks on the Stanford Achievement Test were examined at the end of first and second grade. Although discontinued RR children were close behind their RS peers at the end of first grade, their mean rankings fell somewhat by the end of second grade. However, they were still ahead of not discontinued RR children and WL children.

New York University

In the New York City (Jaggar & Smith-Burke, 1995) and New York State (Jaggar & Simic, 1996) follow-up studies, researchers conducted assessments on second and third graders who had been discontinued from the RR program. The studies follow a design very similar to that used in the Columbus follow-up studies previously described (Pinnell, DeFord, & Lyons, 1988; Ohio State University, 1989). Successfully discontinued RR children were tested at the end of second and third grades with the text reading measure. For the NYC studies, the text reading measure included levels 1 - 30; for the NY State studies, the ceiling was raised to 34. The NY State studies also included the Slosson Word Test, an individually-administered assessment in which the child is asked to read words from a context-free list. Lists increase with difficulty as grade levels increase.

The NY State studies are notable for their size. One thousand, five hundred, ninety-six (1596) second graders and 604 third graders were included. No attempt was made to select a smaller sample for testing, so as much of the entire population as could be found was tested. The NYC studies were considerably smaller (3 cohorts, the smallest of which was 48 children, and the largest of which was 355 children) because the population of RR children was smaller.

In both the NYC and NY State studies, as in the Columbus, Ohio, studies, a new RS was drawn every year. As described previously, this has limitations. Other limitations of these studies include ceiling effects on the dependent measures and the fact that only data from successfully discontinued RR children were included.

Both studies concluded that discontinued RR children's gains are sustained. Discontinued RR children from NYC were within the average band at the end of second and third grades.

Discontinued RR children from NY State were not significantly different from RS children at the ends of grades 2, 3, or 4 on either measure of literacy skill.

Wake County, North Carolina

In a coordinated research effort out of Raleigh, North Carolina (Wake County Public Schools System, 1995) two studies were done, one by RR staff and one by staff of the school district evaluation and research center. In the RR staff study, control students were 1-1 matched with RR students by first grade pretest scores. Successfully discontinued RR children were compared to matched controls on the California Achievement Test (CAT). They found that 32% of D RR and 27% of comparison children performed above 50th percentile on the CAT in the spring of third grade.

In the study done by staff at the evaluation and research office, RR and control children were alternately assigned to groups (either RR or another intervention). RR children's and control children's scores were compared to state means on the End-Of-Grade reading test (EOG, a state standardized test).

One problem with this study is that it relied on state means. No information was given about where the not-at-risk children in the school system were scoring. No information was given about how many students in the system (non-RR) were reading on grade level.

No significant differences were found between control and RR, even when only successfully discontinued RR were counted. Based on these findings, the report concludes that RR is not cost effective for the district.

In order to reconcile this conclusion with the opposite conclusion reached by other researchers, it is important to note the different measures used as well as the fact that the Wake

County Public Schools were serving only 4.4% of their students at the time of the report. The RR program estimates that, in most classrooms, 20% of children will fall into the at-risk category. Therefore, it seems plausible that the data from the 15.6% of the children who might have participated in RR but did not could have contributed to a different result. This is not to say that the children in the lowest 4.4% of the classroom will not meet with success though RR, only that a smaller proportion of the lowest 4.4% than of the lowest 20% will probably achieve success through RR.

Battelle

Ohio State University, the national headquarters for RR in the United States, contracted an outside agency to conduct an unbiased, statewide (Ohio), longitudinal assessment of the program. The study was conducted by Battelle (1995).

Effects of type of reading program, school district, school, teacher, and student were all considered in the design. Sampling was done by region to ensure geographic (and socioeconomic, and pedagogic) diversity. The study was large, involving 887 RR children and 1078 children receiving other services. The sample included children from rural, small town, and urban schools, from all across Ohio. All children in RR, regardless of outcome, as long as they had had at least 20 days of service, were included as RR children.

Teachers were asked to assess what grade level the student had achieved by the end of the year. In addition, a standardized test of reading comprehension was used, the MAT6. Baseline data were also collected at the beginning of year 1 using the MAT6. This baseline was used as a covariate in the ANCOVA analyses. One of the problems with the design concerns this covariate. Most at-risk children would score zero at the beginning of first grade on a test of reading

comprehension, although their skills might differ considerably. In other words, the measure used as a covariate had a large floor effect.

General findings were that RR children scored higher than children receiving other reading services at the end of year 1, but that the two groups were not different at the end of the second, third, and fourth grades. RR children who had discontinued from the program sustained more gains through the fourth grade than those who had not discontinued. No differences were found between children who had had preschool and those who had not, nor between genders. No differences were found by region.

Westbrook, Maine

Although not technically a longitudinal study, information was systematically gathered on an annual basis from one of the first districts in Maine to implement RR (Jackman, Unpublished). The Teacher Leader there collected data on the number of resource room students, number of retentions, and average scores on the MEA reading test, a statewide achievement test administered in the 4th grade in all Maine public schools. The data have been used locally to examine the effects of the RR program and to inform educational decisions.

The data gathered are at the school level. Specifically, they indicate a large decrease in the number of resource room students (an alternate intervention for children who have not been successful with only classroom instruction) and retentions after RR was implemented in the district. In addition, the average MEA (Maine Educational Association test, a statewide assessment administered in 4th grade) reading score in the district increased when the second cohort of children whose low-achieving peers had recieved RR in first grade took the test.

Texas Women's University

Two Texas researchers conducted and publised the results from a longitudinal study of RR (Askew & Frasier, 1997). In the study, discontinued RR children were compared to their RS peers. The researchers measured text reading level, retelling ability, fluency of oral reading, hearing and recording sounds, and spelling. All measures were part of or derived from the Observation Survey (Clay, 1993). The text reading measure was scored as described previously. RR teacher leaders assigned children scores for their retellings and fluency, after listening to them read their text reading passage.

Hearing and recording sounds (HRS) is another component of the Observation Survey which assesses a child's ability to break words and sentences down into sounds. Typically, any phonetically correct representation earns full credit, regardless of spelling convention. The spelling test in this study was derived from the HRS test, with the exception that only conventional spellings of words earned points.

Testing was done by RR teacher leaders at the end of their training year or by experienced teacher leaders in the field. Scoring was done "blind," without knowledge of whether a child was RR or RS.

RR children scored within an "expanded average band" of RS children (the RS mean, plus or minus one whole SD) on text reading, hearing and recording sounds, and spelling. On three measures of fluency (phrasing, smoothness, and pacing), and three measures of retelling (text-based, prior knowledge, and language/organization), RR children scored, on average, within 1 SD of RS children, on average. No significant differences were found between RR and RS children

on any measure except one component of fluency. RS children were significantly faster readers than RR children.

Teacher perceptions of children's abilities and predictions of future literacy success put most RR children in the average range; more RS were in the higher ability range. The RR children displayed more "reading work," of repetitions, self-corrections, and multiple attempts.

There were two major limitations to the study. First, only discontinued RR children were included. Second, the measures used suffered from ceiling effects. Both of these limitations have been discussed in relation to previous studies. One powerful piece of evidence that ceiling effects were a problem is that the high end of the "average band" exceeded the maximum possible score for three of the measures. The authors also note that text scores were "extraordinarily high for both groups" (page 32). They suggest that future studies should look into other measures, to possibly include standardized assessments, including those that ask children to read silently and then answer questions.

The University of Melborne

Rowe (1997) conducted a major longitudinal study of reading acquisition in Australia, which included participation in the RR program as one of the variables. Over 5,000 students, aged 5-14, from grades 2 through 6, were included. This large-scale, longitudinal study, conducted in Australia, addressed the factors (student, teacher, and school) that contribute to good reading performance.

Five cohorts of students (initially at grades 3, 5, 7, & 9) were measured over a three year period, for data that spanned grades 1 through 11. Five thousand, six hundred (5600) students were selected from 280 schools. Socio-economic indicators were measured, including mother's

education, father's education, and mother's and father's occupational classification. Students' reading activity at home and attitudes toward reading were measured through self-report. Attitudes toward reading were measured with three Likert-scale items: *Reading is useful, I read well,* and *I enjoy reading*. Students' reading achievement was measured using both scores on a standardized reading comprehension test and teacher ratings.

The findings clearly showed that the range of students' reading ability grows as children advance through the grades, with a huge difference between the 10th and 90th percentiles by the 11th grade. Attentiveness in school accounted for more variance in reading achievement than any other student variable (around 20%). Reading activity at home was a close second, at around 15% for 7 years and up (less for younger children). Attitudes about reading accounted for 5 - 10% of the variance in reading achievement. Socio-economic status was an unimportant factor.

Students who had been in RR showed a significantly smaller variance in reading achievement than their peers. Although they were slightly behind their non-RR peers, they were very close. The author notes that "those students who had been identified as readers at risk and placed in a RR program have benefitted notably from participation" (Rowe, 1997, p. 76).

Future Research

Future research into the long term impact of the RR program is needed. Many of the studies reviewed here suffered from serious methodological flaws. Ceiling effects were a problem for studies that relied on the text reading level measure beyond grade 1. Another common problem was including only discontinued RR children. Finally, implementation levels differ so much from school to school, yet this information is rarely reported. A longitudinal study is

planned for Maine. Data collection on second graders is scheduled to begin in the spring of 1998. Important features of this study will be:

- Valid, reliable assessments that do not suffer ceiling or floor effects.
- Including all RR children in the data set.
- Keeping the same children in the data set year after year (true longitudinal design).
- Large enough sample to draw conclusions, small enough for good data collection.
- Both RS and WL children used as comparison groups.
- Noting implementation level when evaluating the program.

References

Askew, B.J. & Frasier, D.F. (1997). Sustained effects of Reading Recovery intervention on the cognitive behaviors of second grade children and the perceptions of their teachers. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 18 - 38). Portsmouth, NH: Heinemann.

Battelle (1995). *Longitudinal Study of Reading Recovery: 1990-91 through 1993-94.* Ohio Department of Education: Technical Report.

Clay, M. M. (1982). *Observing young readers.* Portsmouth, NH: Heinemann.

Clay, M. M. (1985). *The early detection of reading difficulties (3rd ed.).* Portsmouth, NH: Heinemann.

Clay, M. M. (1991). *Becoming literate: The construction of inner control.* Portsmouth, NH: Heinemann Educational Books, Inc.

Clay, M. M. (1993). *An observation survey of early literacy achievement.* Portsmouth, NH: Heinemann Educational Books, Inc.

Curtin, Josephine. (1993). *The Effectiveness of the Reading Recovery Program on Reading Achievement.* Chicago, IL: Chicago Public Schools. (ERIC Document Reproduction Service No. ED 363 863)

Jackman, P. (Unpublished). Unpublished data from the Westbrook, Maine, school district. Available to the author through personnal communication.

Jaggar, A. M. & Simic, O. (1996.) A *Four-Year Follow-Up Study of Reading Recovery Children in New York State: Preliminary Report.* New York: New York University School of Education. (Technical Report).

Jaggar, A. M. & Smith-Burke, M. T. (1995). *The Reading Recovery Program in New York City, A Five-Year Summary Report, 1989-1994.* New York: New York University School of Education (Technical Report)

Ramaswami, Soundaram. (1994). *The Differential Impact of Reading Recovery on the Achievement of First Graders in the Newark School District, 1991-1993.* Newark, NJ: Newark Board of Education, Office of Planning, Evaluation, and Testing. (ERIC Document Reproduction Service No. ED 374 180)

Rowe, K. J. (1997). Factors affecting students' progress in reading: Key findings from a longitudinal study. In S. L. Swartz & A. F. Klein (Eds.), *Research in Reading Recovery* (pp. 18 - 38). Portsmouth, NH: Heinemann.

Wake County Public Schools System. (1995). *WCPSS Reading Recovery 1990-1994.* (Evaluation Report No. 95.09A). Raleigh, NC: WCPSS Department of Evaluation and Research. (ERIC Document Reproduction Service No. ED 383 733)